

**STUDENTŲ ĮTRAUKIMO Į MOKSLINĘ VEIKLĄ SKATININAMOJO KONKURSO TEMA**

<p>Temos pavadinimas: Teksto generavimo aptikimas grįstas mašininio mokymo algoritmais</p>
<p>Tikslas: sukurti ir ištirti mašininio mokymo modelius, kurie leis aptikti ar studentas naudojo teksto generavimą atsakant į tarpinio atsiskaitymo klausimus.</p>
<p>Trumpas temos vykdymo aprašymas (ne daugiau kaip 2000 ženklų): Sukūrus didžiuosius kalbos modelius tokius kaip chatGPT, Copilot ir kitus, studentai pradėjo sukčiauti atlikdami įvairaus tipo užduotis. Natūralų, jog aptikti sukčiavimą yra sudėtinga užduotis, kadangi panaudojus kelis teksto generatorius ar perfrazavus savo nuomone pateiktas tekstas gali būti neatpažintas. Todėl šio tyrimo metu yra siekiama išbandyti mašininio mokymo modelių tinkamumą atpažįstant žmogaus parašytą tekstą nuo generuoto. Atliekant šį tyrimą reikės atlikti šiuos žingsnius:</p> <ol style="list-style-type: none"><li>1. Pavasario semestre kartu su dėstytoju surinkti naują duomenų aibę atliekant tarpinį atsiskaitymą kurse „Mašininis mokymas“.</li><li>2. Sužymėti duomenis.</li><li>3. Pasirinkti mašininio mokymo algoritmus ir juos panaudojus sukurti mašininio mokymo modelius.</li><li>4. Ištirti mašininio mokymo modelių efektyvumą.</li><li>5. Aprašyti tyrimo rezultatus.</li></ol>
<p>Temą siūlantis mokslininkas/dėstytojas: doc. dr. Pavel Stefanovič</p>

## THE TOPIC OF A COMPETITION PROMOTING STUDENT ENGAGEMENT IN SCIENTIFIC ACTIVITIES

Topic: Text generation detection based on machine learning algorithms

Goal: to develop and investigate machine learning models that will detect whether a student has used text generation tools to answer midterm questions.

Short description:

After the creation of large language models like chatGPT, Copilot and others, students started cheating on different types of tasks. It is natural that detecting cheating is a difficult task, since using several text generators or paraphrasing your own opinion may not recognize the text presented. Therefore, this research aims to test the suitability of machine learning models in recognizing human-written text from generated text. The following steps will be required for this investigation:

1. In the spring semester, together with the supervisor, collect a new set of data during the midterm assessment in the course of "Machine Learning".
2. Label the dataset.
3. Select machine learning algorithms and use it for machine learning models.
4. Investigate the effectiveness of machine learning models.
5. Describe the results of the study.

Supervisor: doc. dr. Pavel Stefanovič